

特開平 10-254350

(43) 公開日 平成 10 年 (1998) 9 月 25 日

(51) Int. Cl. <sup>6</sup>	識別記号	庁内整理番号	F I	技術表示箇所
G09B 19/04			G09B 19/04	
G10L 3/00	531		G10L 3/00	531 K
	551			551 Z

審査請求 未請求 請求項の数 3 O L (全 6 頁)

(21) 出願番号 特願平 9-59138

(22) 出願日 平成 9 年 (1997) 3 月 13 日

(71) 出願人 000006013

三菱電機株式会社

東京都千代田区丸の内二丁目 2 番 3 号

(72) 発明者 鈴木 忠

東京都千代田区丸の内二丁目 2 番 3 号 三

菱電機株式会社内

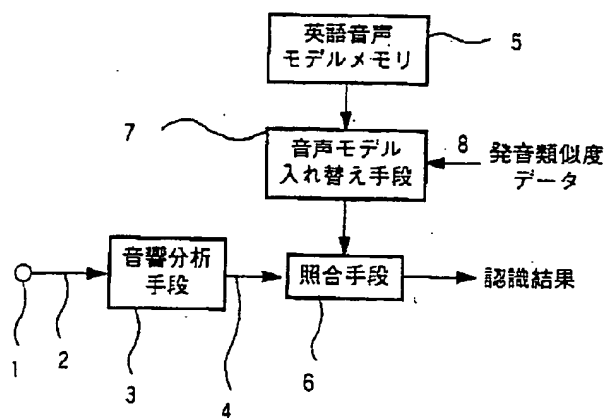
(74) 代理人 弁理士 宮田 金雄 (外 2 名)

(54) 【発明の名称】 音声認識装置

(57) 【要約】

【課題】 本発明は音声認識装置において、任意の外国語について、これを母国語としない話者すなわち非母国語話者の発声に特有の発音の曖昧さや誤りがあっても、認識精度を向上する。

【解決手段】 外国語音声モデルメモリに格納されている音声モデルを読み出し、発音類似度データに従って音声モデルの入れ替えを行って出力することにより、非母国語話者特有の発音の曖昧さや誤りに合せた音声モデルによる認識照合処理が実行でき、かくして、非母国語話者の発声に特有の発音の曖昧さや誤りがあっても、認識精度を向上できる。



## 【特許請求の範囲】

【請求項1】 入力される音声信号に対し音響分析を施し、特徴ベクトル時系列を出力する音響分析手段と、外国語の音声モデルを記憶する外国語音声モデルメモリと、前記外国語音声モデルメモリ内の音声モデルを、入力される発音類似度データにしたがって入れ替えて出力する音声モデル入れ替え手段と、前記音声モデル入れ替え手段から出力される音声モデルを用いて、上記音響分析手段に入力された未知入力信号に対する特徴ベクトル時系列の認識照合処理を行い、認識結果を出力する照合手段とを備えることを特徴とする音声認識装置。

【請求項2】 前記音声モデル入れ替え手段において使用される発音類似度データを作成する発音類似度データ作成手段として、前記音響分析手段から出力される特徴ベクトル時系列に対して、音声モデルとの対応付けを行って出力する音声モデル対応付け手段と、この音声モデル対応付け手段の出力を格納する特徴ベクトルメモリと、前記外国語音声モデルメモリに格納されている音声モデルと、前記特徴ベクトルメモリに格納されている音声モデルとの対応付け情報が付与された特徴ベクトル時系列との尤度演算を行い、その結果を発音類似度データとして出力する尤度演算手段とを備えることを特徴とする請求項1に記載の音声認識装置。

【請求項3】 前記音声モデル入れ替え手段において使用される発音類似度データを作成する発音類似度データ作成手段として、前記音響分析手段から出力される特徴ベクトル時系列を用いて音声モデルの学習を行い出力する音声モデル学習手段と、前記音声モデル学習手段の出力である音声モデルデータを格納する音声モデルメモリと、前記音声モデル学習手段により得られた音声モデルと前記外国語音声モデルメモリ内に格納されている音声モデルとを用いて、音声モデル同士の類似を計算し、これを発音類似度データとして出力する類似度演算手段とを備えることを特徴とする請求項1記載の音声認識装置。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】 本発明は、任意の外国語について、これを母国語としない話者が発声した該外国語の音声認識の対象とする音声認識装置に関するものである。

## 【0002】

【従来の技術】 従来、共通語としての英語により対話が行われる航空機の操縦や管制業務等においては、日本人も英語による対話を行う必要がある。このような業務において音声認識システムの適用を図る場合、英語音声の認識装置が必要になる。

【0003】 図4は文献「音声認識の基礎（下）」(Lawrence Rabiner, Bing-Hwang Juang共著、古井貞▲おき▼監訳、NTTアドバンステクノロジー株式会社、247

頁～273頁、1995年11月発行)に示されているサブワード音声単位による英語音声認識装置の構成の一例である。図において、3は入力端1から入力される音声信号2に対し音響分析を行い特徴ベクトル時系列4を出力する音響分析手段、5はサブワード音声単位を表す音声モデルを格納する英語音声モデルメモリ、6は特徴ベクトル時系列4に対し英語音声モデルメモリ5に格納されている音声モデルを用いて認識照合処理を行い認識結果を出力する照合手段である。

【0004】 次に動作について説明する。サブワード音声単位としては前記文献の250頁に示された疑似音素単位を用いる場合を例に取る。疑似音素単位は、英語音声に依存して決定され、英語にあらわれる種々の発音を網羅するように設計されている。各疑似音素単位をあらわす音声モデルは、予め大量の英語母国語話者音声データから学習され、英語音声モデルメモリ5に格納されている。また照合手段6には、認識対象語彙である複数の単語や文章が、音声モデルすなわち疑似音素単位のラベルの系列として記憶されている。このラベル系列も音声モデルと同様に英語母国語話者の発声様態に合わせて構成されている。

【0005】 入力端1から入力された音声信号2は音響分析手段3において、分析フレーム毎に音響分析処理され、特徴ベクトル時系列4として出力される。照合手段6では、英語音声モデルメモリ5に格納されている音声モデルを入力として、前記音響分析手段3の出力である特徴ベクトル時系列4に対する認識照合処理を行う。この認識照合処理は、認識対象語彙のラベル系列の各ラベルに対応する音声モデルと特徴ベクトル時系列4との尤度を求めることでラベル系列全体での尤度を計算し、この尤度が最大となる単語もしくは文章を認識結果として出力する。

## 【0006】

【発明が解決しようとする課題】 ところで英語母国語話者のような発音ができなくとも会話が成立することは、多くの人が経験的に知っており、業務対話なども同様である。このような業務分野への音声認識システムの適用を図る場合、業務遂行上問題が発生しない程度の日本人なまりを含む英語音声の高精度認識が必要となる。ところが、従来の英語音声認識装置は上記のように、英語を母国語とする話者の音声基準として設計され学習されているため、英語を母国語としない日本人が発声した英語入力音声に対しては、日本人特有のなまりの影響により、十分な認識性能が得られにくいという問題があった。これは英語と日本人の組み合わせに限らず、任意の外国語とそれを母国語としない話者の全ての組み合わせについて共通の問題である。

【0007】 本発明は上記の問題を解決するためになされたもので、任意の外国語について、これを母国語としない話者すなわち非母国語話者の発声に特有の発音の曖

味さや誤りがあっても、認識精度を向上し得る音声認識装置を提供しようとするものである。

【0008】

【課題を解決するための手段】この発明に係る音声認識装置は、入力される音声信号に対し音響分析を施し、特徴ベクトル時系列を出力する音響分析手段と、外国語の音声モデルを記憶する外国語音声モデルメモリと、外国語音声モデルメモリ内の音声モデルを、入力される発音類似度データにしたがって入れ替えて出力する音声モデル入れ替え手段と、音声モデル入れ替え手段から出力される音声モデルを用いて、音響分析手段に入力された未知入力信号に対する特徴ベクトル時系列の認識照合処理を行い、認識結果を出力する照合手段とを備えるものである。

【0009】さらに次の発明に係る音声認識装置は、これに加えて、音声モデル入れ替え手段において使用される発音類似度データを作成する発音類似度データ作成手段として、音響分析手段から出力される特徴ベクトル時系列に対して、音声モデルとの対応付けを行って出力する音声モデル対応付け手段と、この音声モデル対応付け手段の出力を格納する特徴ベクトルメモリと、外国語音声モデルメモリに格納されている音声モデルと、特徴ベクトルメモリに格納されている音声モデルとの対応付け情報が付与された特徴ベクトル時系列との尤度演算を行い、その結果を発音類似度データとして出力する尤度演算手段とを備えるものである。

【0010】さらにまた次の発明に係る音声認識装置は、これに加えて、音声モデル入れ替え手段において使用される発音類似度データを作成する発音類似度データ作成手段として、音響分析手段から出力される特徴ベクトル時系列を用いて音声モデルの学習を行い出力する音声モデル学習手段と、音声モデル学習手段の出力である音声モデルデータを格納する音声モデルメモリと、音声モデル学習手段により得られた音声モデルと外国語音声モデルメモリ内に格納されている音声モデルとを用いて、音声モデル同士の類似を計算し、これが発音類似度データとして出力する類似度演算手段とを備えるものである。

【0011】

【発明の実施の形態】以下図面について、本発明の実施の形態を説明する。

【0012】実施の形態1. 図4との対応部分に同一符号を付けた図1は、本発明による音声認識装置の実施の形態1の構成を示す。図において、7は入力される発音類似度データ8に従って、英語音声モデルメモリ5に格納されている音声モデルを読み出し、音声モデルの入れ替えを行って出力する音声モデル入れ替え手段である。

【0013】次に動作について説明する。図4と同様に、予め疑似音素単位を表す音声モデルが、大量の英語母国語話者音声データから学習され、英語音声モデルメ

モリ5に格納されている。また、照合手段6にも認識対象語彙である複数の単語や文章が、音素モデルのラベル系列として記憶されている。このラベル系列や音声モデルが英語母国語話者の発声の形態に合わせて構成されている点も図4の音声認識装置と同様である。

【0014】発音類似度データ8は、日本人が発声した英語音声におけるラベルiの疑似音素単位が、英語母国語話者の音声におけるラベルjの疑似音素単位とどの程度類似しているかを表す類似度 $L_{ij}$ で表されている。ただし、 $1 \leq i \leq N$ 、 $1 \leq j \leq N$ 、Nは疑似音素単位の種類数である。音声モデル入れ替え手段7は、英語音声モデルメモリ5に格納されている音声モデルについて、ラベルiの音声モデル $M_i$ の代わりに、類似度 $L_{ij}$ を $1 \leq j \leq N$ において最大にする $j_{max}$ のラベルを持つ音声モデル $M_{j_{max}}$ を出力することで、音声モデルの入れ替え処理を行う。

【0015】これにより、ラベルiの音声モデルとして、日本人が発声した英語音声におけるラベルiの疑似音素単位に一番近い英語母国語話者の音声における疑似音素単位の音声モデルが使われることになり、日本人特有の発音上の誤りを表現した音声モデルを用いて、認識を行うことが可能になる。

【0016】なお音声モデルが表すサブワードは、疑似音素単位に限定されるものではない。また対象となる言語と話者は英語ならびに日本人に限定されるものではなく、他の外国語ならびに非母国語話者であれば、上述の実施の形態と同様の効果を実現できる。

【0017】実施の形態2. 図1との対応部分に同一符号を付けた図2は、本発明による音声認識装置の実施の形態2の構成を示し、その特徴として発音類似度作成手段を有する。図において、9は音響分析手段3の出力である特徴ベクトル時系列4の出力先を切り替えるスイッチである。また10はスイッチ9を介して音響分析手段3の出力である特徴ベクトル時系列4を入力として、この特徴ベクトル時系列に対し疑似音素単位を表す音声モデルの対応付けを行い、特徴ベクトル時系列にこの対応付け情報を付与して出力する音声モデル対応付け手段である。

【0018】さらに11は、音声モデル対応付け手段の出力である音声モデルとの対応付け情報を付与された特徴ベクトル時系列を格納する特徴ベクトルメモリである。さらにまた12は、英語音声モデルメモリに格納されている音声モデルと特徴ベクトルメモリ11に格納されている特徴ベクトル時系列との尤度演算を行い、その結果を発音類似度データ8として出力する尤度演算手段である。

【0019】次にこの音声認識装置において、発音類似度データを作成する動作を説明する。まず発音類似度データを作成する際にはスイッチ9はa側に押され、音響分析部3の出力は音声モデル対応付け手段10に流れる

10

20

30

40

50

ようにセットされる。発音類似度データを作成するため、日本人が発声した内容既知の英語音声が入力端1より入力され、音響分析部3において音響分析され特徴ベクトル時系列4として出力される。

【0020】音声モデル対応付け手段10では、音響分析手段3から出力された日本人英語音声の特徴ベクトル時系列に対し、この日本人英語音声の発声内容に沿った音声モデルの系列を用意し、尤度最大基準で、各特徴ベクトルに対する音声モデルの対応付けを決定する。このときに用いる音声モデルには、英語音声モデルメモリ内の音声モデルを用いても良いし、英語音声モデルメモリ内の音声モデルを用いて得られた対応付けに従って、学習し直した音声モデルを用いても良い。また、この対応付けと音声モデルの再学習を複数回繰り返してから、最終的な対応付けを決定してもかまわない。音声モデルとの対応付け情報を付与した特徴ベクトル時系列は特徴ベクトルメモリ11に格納される。

【0021】尤度演算手段12は、特徴ベクトルメモリ11に格納されている音声モデルとの対応付け情報つきの特徴ベクトル時系列の中から、ラベルiの音声モデルと対応づけられた複数個の部分特徴ベクトル時系列を取り出し、英語音声モデルメモリ内のラベルjの音声モデルとの平均尤度を求め、これを日本人英語音声におけるラベルiの疑似音素単位に対する、英語母国語話者のラベルjの疑似音素単位の類似度 $L_{ij}$ として出力する。これを、 $1 \leq i \leq N$ および $1 \leq j \leq N$ （ただしNは疑似音素単位の数）の範囲で行い、発音類似度データ8として出力する。

【0022】以上の動作で作成された発音類似度データを用いて、図1について上述したと同様の日本人英語音声認識を行う。この認識時にはスイッチ9はb側に押され、音響分析手段3の出力である特徴ベクトル時系列4は照合手段6に入力される。これにより、実際に日本人が発声した英語音声データに従って、発音類似度データが決定されるので、日本人英語の認識精度を一段と向上できる。

【0023】なお音声モデルが表すサブワードは、疑似音素単位に限定されるものではない。また対象となる言語と話者は英語ならびに日本人に限定されるものではなく、他の外国語ならびに非母国語話者であれば、上述の実施の形態と同様の効果を実現できる。

【0024】実施の形態3. 図1及び図2との対応部分に同一符号を付けた図3は、本発明による音声認識装置の実施の形態3の構成を示し、その特徴として発音類似度作成手段を有する。図において、13はスイッチ9を介して音響分析手段3の出力である特徴ベクトル時系列4を入力し、この特徴ベクトル時系列を用いて日本人英語音声における音声モデルの学習を行う音声モデル学習手段である。また14は、音声モデル学習手段の出力である日本人英語の音声モデルを格納する音声モデルメモ

リである。さらに15は、英語音声モデルメモリに格納されている音声モデルと音声モデルメモリ14に格納されている日本人英語の音声モデルとの類似度演算を行い、その結果を発音類似度データ8として出力する尤度演算手段である。

【0025】次にこの音声認識装置において、発音類似度データを作成する動作を説明する。まず発音類似度データを作成する際にはスイッチ9はa側に押され、音響分析部3の出力は音声モデル学習手段13に流れるようにセットされる。発音類似度データを作成するため、日本人が発声した内容既知の英語音声が入力端1より入力され、音響分析部3において音響分析され特徴ベクトル時系列4として出力される。

【0026】音声モデル学習手段13では、音響分析手段3から出力された日本人英語音声の特徴ベクトル時系列と、この日本人英語音声の発声内容に沿った音声モデルとの照合と、その結果得られる特徴ベクトルと音声モデルの対応付けによる音声モデルの更新の繰り返しにより、日本人英語音声における疑似音素単位を表す音声モデルの学習を行い、得られた音声モデルを音声モデルメモリに格納する。

【0027】類似度演算手段15は、音声モデルメモリ14に格納されている日本人英語のラベルiの音声モデルと、英語音声モデルメモリ内のラベルjの音声モデルとの類似度を計算し、これを日本人英語音声におけるラベルiの疑似音素単位に対する、英語母国語話者のラベルjの疑似音素単位の類似度 $L_{ij}$ として出力する。これを、 $1 \leq i \leq N$ および $1 \leq j \leq N$ （ただしNは疑似音素単位の数）の範囲で行い、発音類似度データ8として出力する。

【0028】類似度 $L_{ij}$ としては、音声モデルが表現する疑似音素単位のスペクトル上での類似度を反映する類似度式ならば何を用いても良い。例えば、音声モデルが単一の音声スペクトルで表されている場合は、パワースペクトル距離の逆数に重みをかけたものや、スペクトルから計算されるケプストラム係数をベクトルとした時のベクトル間のユークリッド距離の逆数などである。音声スペクトルの時系列で表される場合も、線形伸縮や非線形伸縮により系列数を合わせて、おなじ時刻のスペクトル間の距離の平均の逆数などが用いられる。また、音声モデルが離散型もしくはセミ連続型のHMM (Hidden Markov Model) ならば、同じ出力ラベルに対する出力確率をかけ、すべての出力ラベルについて総和した値でもかまわない。連続分布型のHMMにおいても、分布間において定義される種々の距離値の逆数などが用いられる。

【0029】以上の動作で作成された発音類似度データを用いて、図1について上述したと同様の日本人英語音声認識を行う。この認識時にはスイッチ9はb側に押され、音響分析手段3の出力である特徴ベクトル時系列4は照合手段6に入力される。これにより、実際に日本人

7

が発声した英語音声データに従って、発音類似度データが決定されるので、日本人英語の認識が精度良く行える。

【0030】なお音声モデルが表すサブワードは、疑似音素単位に限定されるものではない。また対象となる言語と話者は英語ならびに日本人に限定されるものではなく、他の外国語ならびに非母国語話者であれば、上述の実施の形態と同様の効果を実現できる。

【0031】

【発明の効果】以上のようにこの発明によれば、外国語音声モデルメモリに格納されている音声モデルを読み出し、発音類似度データに従って音声モデルの入れ替えを行って出力することにより、非母国語話者特有の発音の曖昧さや誤りに合わせた音声モデルによる認識照合処理が実行でき、かくして、非母国語話者の発声に特有の発音の曖昧さや誤りがあっても、認識精度を向上し得る音声認識装置を実現できる。

【0032】また次の発明によれば、非母国語話者が発声した音声データの特徴ベクトル時系列に対する音声モデルの対応付けを求め、この対応付け情報を付与した特徴ベクトル時系列を出力し、音声モデルとの対応付け情報を持つ特徴ベクトル時系列を入力として、外国語音声モデルメモリ上の音声モデルとの尤度演算により、発音類似度データを作成し出力することにより、非母国語話者特有の発音の曖昧さや誤りに合わせた音声モデルによる認識照合処理が実行され、さらに認識精度を向上し得る音声認識装置を実現できる。

【0033】またさらに次の発明によれば、非母国語話者が発声した音声データの特徴ベクトル時系列を用いて音声モデルの学習を行い、その音声モデルと外国語音声

8

モデルメモリ内の音声モデルとの類似度を計算し、これを発音類似度データとして出力することにより、非母国語話者特有の発音の曖昧さや誤りに合わせた音声モデルによる認識照合処理が実行され、さらに認識精度を向上し得る音声認識装置を実現できる。

【図面の簡単な説明】

【図1】 この発明による音声認識装置の実施の形態1の構成を示すブロック図である。

【図2】 この発明による音声認識装置の実施の形態2の構成を示すブロック図である。

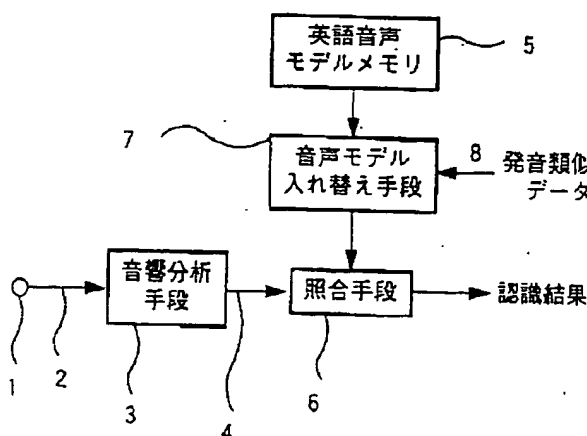
【図3】 この発明による音声認識装置の実施の形態3の構成を示すブロック図である。

【図4】 従来の音声認識装置の構成を示すブロック図である。

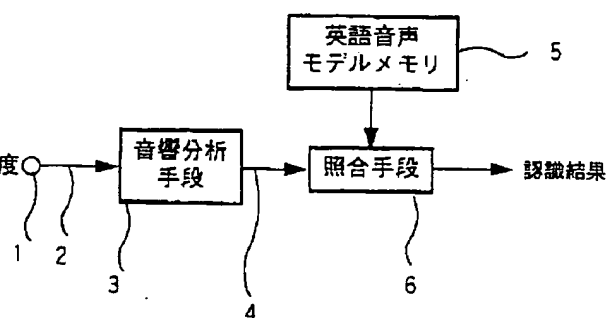
【符号の説明】

- 1 入力端
- 2 入力音声信号
- 3 音響分析手段
- 4 特徴ベクトル時系列
- 5 英語音声モデルメモリ
- 6 照合手段
- 7 音声モデル入れ替え手段
- 8 発音類似度データ
- 9 スイッチ
- 10 音声モデル対応付け手段
- 11 特徴ベクトルメモリ
- 12 尤度演算手段
- 13 音声モデル学習手段
- 14 音声モデルメモリ
- 15 類似度演算手段

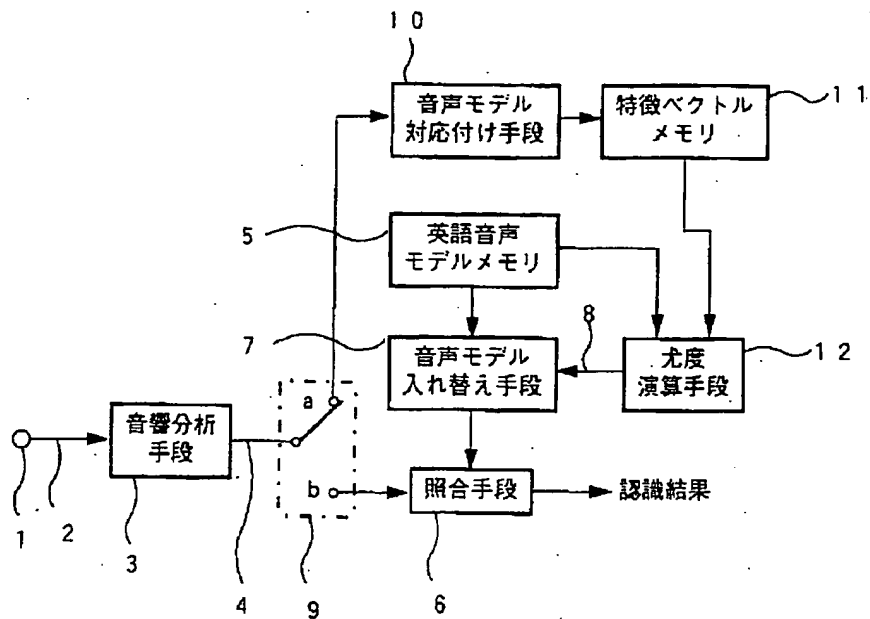
【図1】



【図4】



【図2】



【図3】

